JAMA | Special Communication

# Reporting of Multi-Arm Parallel-Group Randomized Trials
## Extension of the CONSORT 2010 Statement

Edmund Juszczak, MSc; Douglas G. Altman, DSc; Sally Hopewell, DPhil; Kenneth Schulz, PhD

**IMPORTANCE** The quality of reporting of randomized clinical trials is suboptimal. In an era in which the need for greater research transparency is paramount, inadequate reporting hinders assessment of the reliability and validity of trial findings. The Consolidated Standards of Reporting Trials (CONSORT) 2010 Statement was developed to improve the reporting of randomized clinical trials, but the primary focus was on parallel-group trials with 2 groups. Multi-arm trials that use a parallel-group design (comparing treatments by concurrently randomizing participants to one of the treatment groups, usually with equal probability) but have 3 or more groups are relatively common. The quality of reporting of multi-arm trials varies substantially, making judgments and interpretation difficult. While the majority of the elements of the CONSORT 2010 Statement apply equally to multi-arm trials, some elements need adaptation, and, in some cases, additional issues need to be clarified.

**OBJECTIVE** To present an extension to the CONSORT 2010 Statement for reporting multi-arm trials to facilitate the reporting of such trials.

**DESIGN** A guideline writing group, which included all authors, formed following the CONSORT group meeting in 2014. The authors met in person and by teleconference bimonthly between 2014 and 2018 to develop and revise the checklist and the accompanying text, with additional discussions by email. A draft manuscript was circulated to the wider CONSORT group of 36 individuals, plus 5 other selected individuals known for their specialist knowledge in clinical trials, for review. Extensive feedback was received from 14 individuals and, after detailed consideration of their comments, a final revised version of the extension was prepared.

**FINDINGS** This CONSORT extension for multi-arm trials expands on 10 items of the CONSORT 2010 checklist and provides examples of good reporting and a rationale for the importance of each extension item. Key recommendations are that multi-arm trials should be identified as such and require clear objectives and hypotheses referring to all of the treatment groups. Primary treatment comparisons should be identified and authors should report the planned and unplanned comparisons resulting from multiple groups completely and transparently. If statistical adjustments for multiplicity are applied, the rationale and method used should be described.

**CONCLUSIONS AND RELEVANCE** This extension of the CONSORT 2010 Statement provides specific guidance for the reporting of multi-arm parallel-group randomized clinical trials and should help provide greater transparency and accuracy in the reporting of such trials.

**Author Affiliations:** NPEU Clinical Trials Unit, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom (Juszczak); Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom (Altman, Hopewell); FHI 360, Durham, North Carolina (Schulz); the University of North Carolina at Chapel Hill, School of Medicine (Schulz).

**Corresponding Author:** Edmund Juszczak, MSc, NPEU Clinical Trials Unit, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Headington, Oxford, OX3 7LF, UK (ed.juszczak@npeu.ox.ac.uk).

Multi-arm randomized clinical trials have several forms but are typically a combination of elements, including multiple active interventions, combinations of active interventions, different doses (or regimens) of an intervention, a placebo, and no active intervention, or treatment as usual. These elements can be combined in various ways resulting in numerous possible trial structures. For example, in a trial with 3 treatment groups, A1 vs A2 vs A3 could represent an evaluation of different doses of the same active intervention. Alternatively, a trial of A1 vs B1 vs C1 could represent an evaluation of 2 different active interventions and a placebo. Moreover, a study comparing A1 vs A2 vs B1 could represent an evaluation of 2 different doses of an active intervention vs another active intervention.

Evaluating more than 1 new intervention concurrently increases the chances of finding an effective intervention.[1] The corresponding increase of efficiency in using a multi-arm (ie, multi-group) design, compared with performing sequential 2-arm (ie, 2 groups) trials, should result in lower cost due to better use of resources. In most cases, sharing a control arm reduces the sample size relative to performing separate 2-arm trials.[2] Offering participants a higher probability of being allocated to a new intervention may result in a greater proportion of eligible individuals enrolling. Some multi-arm trials in oncology have recruited more quickly than comparable 2-arm trials.[1] The argument against multi-arm trials primarily involves statistical power because published trials can have inadequate sample sizes.[3] Given a finite number of potential participants, adding additional treatment groups can further dilute already insufficient power.

Multi-arm trials are relatively common. A detailed review of all randomized trials indexed in PubMed published in 1 month in 2012 showed that 1062 of 1351 (79%) were parallel-group trials[4]; of these 1062 trials, 149 (14%) had 3 groups and 76 (7%) had 4 or more groups.

In this Special Communication, an extension of the Consolidated Standards of Reporting Trials (CONSORT) checklist for the reporting of multi-arm trials is presented, based on the CONSORT 2010 Statement.[5,6] Illustrative examples and explanations for items that differ from the main CONSORT checklist are included. A multi-arm trial is defined as a randomized clinical trial that uses a parallel-group design but has 3 or more groups. For describing the intervention groups in clinical trials, the terms "arms" or "groups" may be used interchangeably, although the term "multi-arm" is used for these reporting guidelines. Other multi-arm or multi-group designs, such as factorial, multi-arm multi-stage, and adaptive, raise rather different issues and are not considered.

## Guidance Development Methods

### Writing Group
The guideline writing group (which included E.J., D.G.A., S.H., and K.S.) formed following a meeting of the CONSORT Group in 2014. The Oxford-based authors met in person with the United States–based author by teleconference bimonthly, and on multiple additional occasions, between 2014 and 2018 to develop and revise the checklist and the accompanying examples and text, with additional discussions by email.

**Key Points**

**Question**  What additional information should be provided when reporting a multi-arm randomized trial that uses a parallel-group design but has 3 or more groups?

**Findings**  This reporting guideline is an extension of the Consolidated Standards of Reporting Trials (CONSORT) 2010 Statement. Ten CONSORT items have been modified, and examples of good reporting and an accompanying explanation for each extension item are provided.

**Meaning**  The guideline checklist can facilitate transparent reporting of multi-arm randomized trials and may help assist evaluations of rigor and reproducibility, enhance understanding of the methodology, and make results more useful for clinicians, journal editors, reviewers, guideline authors, and funders.

### Search Strategy
To identify articles relevant to the methodology of multi-arm randomized trials, a search of PubMed was conducted using the terms "multiarm," "multi-arm," "multiple arm," "multiple treatment," and "multiplicity" combined with the Publication Type term "randomized controlled trial" as a topic, which identified 247 potential articles. One author (S.H.) assessed the titles and abstracts for relevance or potential relevance to this CONSORT extension. The search was supplemented with relevant articles from the personal collections of the authors and by searching the table of contents of books relevant to the methodology of clinical trials for information specific to the conduct and reporting of multi-arm trials.

### Review and Refinement
No formal Delphi process was used in developing this CONSORT extension checklist. The draft manuscript was circulated in April 2017 for review to the wider CONSORT Group, which included 36 individuals, plus 5 selected individuals known for their specialist knowledge in clinical trials. Feedback was received from 14 individuals and, after detailed consideration of their comments, a final revised version of the extension checklist and accompanying explanation was prepared.

## Results

### Checklist Items and Explanation
The **Table** shows the modified checklist for the reporting of multi-arm parallel-group randomized trials; some items are extended to cover the reporting requirements related to the multi-arm design, acknowledging the added complexity imposed by this design. Items that required an extension from the CONSORT 2010 Statement are explained, with illustrative examples of good reporting. For items not mentioned, the advice is the same as for 2-group, parallel randomized trials.

Because all examples have been taken from published articles, it is inevitable that several do not display all of the desirable elements of good reporting. When this is the case, or when there might be ambiguity, the specific aspects of good reporting that are addressed are identified. In some examples, text has been added in brackets to explain the context. The CONSORT 2010 checklist for

Table. Checklist for Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement[a]

| Section/Topic | Item No. | CONSORT 2010 Statement Checklist Item | Multi-Arm Trial Extension |
|---|---|---|---|
| **Title and abstract** | | | |
| | 1a | Identification as a randomized trial in the title | Identification as a multi-arm randomized trial in the title or an indication of the number of treatment groups that the participants were randomly assigned to |
| | 1b | Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)[7] | Specification of the number of treatment groups; details of any groups added or dropped |
| **Introduction** | | | |
| Background and objectives | 2a | Scientific background and explanation of rationale | Rationale for using a multi-arm design |
| | 2b | Specific objectives or hypotheses | Specification of the research question referring to all of the treatment groups Clear statement of all hypotheses to be tested and the primary comparisons involved |
| **Methods** | | | |
| Trial design | 3a | Description of trial design (such as parallel, factorial) including allocation ratio | Specification of the number of treatment groups |
| | 3b | Important changes to methods after trial commencement (such as eligibility criteria), with reasons | Details of any treatment groups added or dropped (if relevant), with reasons, and/or changes to the allocation ratio |
| Participants | 4a | Eligibility criteria for participants | |
| | 4b | Settings and locations where the data were collected | |
| Interventions | 5 | The interventions for each group with sufficient details to allow replication, including how and when they were actually administered | |
| Outcomes | 6a | Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed | |
| | 6b | Any changes to trial outcomes after the trial commenced, with reasons | |
| Sample size | 7a | How sample size was determined | Planned sample size with details of how it was determined for each primary comparison |
| | 7b | When applicable, explanation of any interim analyses and stopping guidelines | |
| **Randomization** | | | |
| Sequence generation | 8a | Method used to generate the random allocation sequence | |
| | 8b | Type of randomization; details of any restriction, such as blocking and block size | |
| Allocation concealment mechanism | 9 | Mechanism used to implement the random allocation sequence, such as sequentially numbered containers, describing any steps taken to conceal the sequence until interventions were assigned | |
| Implementation | 10 | Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions | |
| Blinding | 11a | If applicable, who was blinded after assignment to interventions (eg, participants, care providers, individuals assessing outcomes) and how | |
| | 11b | If relevant, description of the similarity of interventions | |
| Statistical methods | 12a | Statistical methods used to compare groups for primary and secondary outcomes | Explicitly state if no adjustments for multiplicity were applied; if adjustments were applied, state the method used |
| | 12b | Methods for additional analyses, such as subgroup analyses and adjusted analyses | |
| **Results** | | | |
| Participant flow (a diagram is strongly recommended) | 13a | For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome | |
| | 13b | For each group, losses and exclusions after randomization with reasons included | |
| Recruitment | 14a | Dates defining the periods of recruitment and follow-up | If periods of recruitment and follow-up are different across treatment groups (eg, groups were added or dropped), the periods of recruitment and follow-up, reason(s) for the differences, and any statistical implications should be described |
| | 14b | Why the trial ended or was stopped | |
| Baseline data | 15 | A table showing baseline demographic and clinical characteristics for each group | |

*(continued)*

Table. Checklist for Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement[a] (continued)

| Section/Topic | Item No. | CONSORT 2010 Statement Checklist Item | Multi-Arm Trial Extension |
|---|---|---|---|
| Numbers analyzed | 16 | For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups | |
| Outcomes and estimation | 17a | For each primary and secondary outcome, results for each group, and the estimated effect size and its precision, such as 95% CI | Results for each prespecified comparison of treatment groups |
| | 17b | For binary outcomes, presentation of both absolute and relative effect sizes is recommended | |
| Ancillary analyses | 18 | Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory | |
| Harms | 19 | All important harms or unintended effects in each group (for specific guidance, see CONSORT for harms)[8] | |
| Discussion | | | |
| Limitations | 20 | Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses | |
| Generalizability | 21 | Generalizability (external validity, applicability) of the trial findings | |
| Interpretation | 22 | Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence | |
| Other information | | | |
| Registration | 23 | Registration number and name of trial registry | |
| Protocol | 24 | Where the full trial protocol can be accessed, if available | |
| Funding | 25 | Sources of funding and other support (such as supply of drugs), role of funders | |

[a] It is strongly recommended that this checklist is read in conjunction with the CONSORT 2010 Statement Explanation and Elaboration[5] for important clarification on the items.

reporting the abstract of a randomized trial was reviewed. No separate checklist for abstracts is proposed, with the 1 proviso that authors report all of the objectives clearly and specify the number of treatment groups.

## CONSORT Checklist Extension for Multi-Arm Trials

### Title and Abstract
Item 1a. CONSORT 2010: Identification as a randomized trial in the title.

Extension for multi-arm trials: Identification as a multi-arm randomized trial in the title or an indication of the number of treatment groups that the participants were randomly assigned to.

*HARMONY 3: 104-week randomized, double-blind, placebo- and active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin.[9]*

*Efficacy of oral risperidone, haloperidol, or placebo for symptoms of delirium among patients in palliative care: a randomized clinical trial.[10]*

*Multiple Sclerosis-Secondary Progressive Multi-Arm Randomisation Trial (MS-SMART): a multiarm phase IIb randomised, double-blind, placebo-controlled clinical trial comparing the efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis.[11]*

*Explanation*: The ability to identify a report of a randomized trial as such in an electronic database depends largely on how the report was indexed. Indexers may not classify a report as a randomized trial if the authors do not explicitly state this information.[12] To help ensure that a study is appropriately indexed and easily identified, authors should use the word "randomized" in the title and indicate the number of arms (treatment groups) that the participants were randomly assigned to. This issue applies to multi-arm trials also. Article titles normally have a restricted word count, and listing some or all of the interventions is cumbersome, so adding the word "multi-arm" (or multi-group) instead would be efficient and informative.

### Background and Objectives
Item 2a. CONSORT 2010: Scientific background and explanation of rationale.

Extension for multi-arm trials: Rationale for using a multi-arm design.

*Many patients do not respond to monotherapy, and combinations of drugs are often recommended despite little evidence. Lithium plus valproate is often recommended after failure of first-line monotherapy. Should this combination have additive pharmacological effects and prove better than monotherapy, it could be an appropriate first-line therapy. We report here on BALANCE (Bipolar Affective disorder: Lithium/ANti-Convulsant Evaluation), a randomized trial that was designed to establish whether lithium plus valproate semisodium is better than monotherapy with either drug alone for prevention of relapse in bipolar I disorder.[13]*

*Explanation*: When a trial compares 2 parallel groups, it is evident that the aim is a comparison of those groups. With 3 or more intervention groups, however, the intended main comparison or comparisons may not be clear. Because each intervention group should be included only if it contributes to a specific research question,

it follows that each arm should contribute to at least 1 preplanned comparison. Authors should justify the use of a multi-arm design and, in the introduction of the article, indicate why they chose to investigate the interventions they studied and which specific comparisons were planned. In a situation, for example, in which 1 of the planned interventions is a combination of 2 active interventions, authors should comment on why they did not perform a factorial trial. Typically, this "incomplete" factorial design might be used in cases in which it would be unethical to withhold active treatment from a group of patients.

### Item 2b. CONSORT 2010: Specific objectives or hypotheses.

Extension for multi-arm trials: Specification of the research question referring to all of the treatment groups. Clear statement of all hypotheses to be tested and the primary comparisons involved.

*Abstract (Objective): To determine efficacy of risperidone or haloperidol relative to placebo in relieving target symptoms of delirium associated with distress among patients receiving palliative care.*

*Introduction: The aim of this study was to determine if risperidone or haloperidol, given in addition to managing precipitants of delirium and providing individualized supportive nursing care, provides additional benefits in reducing target symptoms of delirium associated with distress when compared with placebo. The primary null hypothesis was that there was no difference between risperidone and placebo, and secondarily, no difference between haloperidol and placebo.[10]*

*Explanation*: Eight possible analyses emanate from a 3-arm trial, (groups A, B, and C) of which most trials will include 2 or 3 (**Box**). The number of potential comparisons proliferates rapidly as the number of intervention groups increases; each group should appear in at least 1 comparison. Thus, unless the intention is only to compare all groups at once (which is not a particularly sensible approach, except for possibly in a dose-response study) there will be at least $k-1$ comparisons made in the analysis of a trial with $k$ treatment arms. The maximum number of 2-group/paired comparisons is $k\times(k-1)/2$ (eg, for a 4-arm trial there are 6 possible 2-group comparisons).

Thus, prespecification of analyses is particularly important, and authors should report all of the planned primary, secondary, and exploratory comparisons. Otherwise, there is a major risk of highlighting and being misled by an observed difference without considering the large number of possible analyses. In all cases, and especially when many comparisons are planned, it is helpful to indicate the primary comparison(s). These comparisons should also be described in the explanation of the planned sample size (item 7a). The planned comparisons may not be considered equally important. For example, one 2-group comparison may be the primary focus of the trial. This distinction is relevant when considering whether to make an adjustment for multiple comparisons. Alternatively, a hierarchical approach to hypothesis testing could prevent any issues with multiple comparisons (item 12a). Some multi-arm trials combine a test of superiority with a test of noninferiority. For example, Foa et al examined whether 10 sessions of prolonged exposure therapy (a trauma-focused cognitive behavioral therapy) delivered over 2 weeks (massed therapy) was more effective than minimal contact (control) and noninferior to 10 sessions of prolonged exposure therapy delivered over 8 weeks (spaced therapy)

for reducing symptom severity among active duty military personnel with posttraumatic stress disorder.[16]

## Methods

### Item 3a. CONSORT 2010: Description of trial design (such as parallel, factorial) including allocation ratio.

Extension for multi-arm trials: Specification of the number of treatment groups.

*In this pragmatic, open-label randomised trial, patients newly diagnosed with Parkinson's disease were randomly assigned (by telephone call to a central office; 1:1:1) between levodopa-sparing therapy (dopamine agonists or MAOBI [monoamine oxidase type B inhibitors]) and levodopa alone.[17]*

*This was a phase 3, randomized, double-blind, placebo- and active-controlled parallel-group study that occurred between 17 February 2009 and 21 March 2013. Eligible patients were stratified by HbA1c level (<8.0% [<63.9 mmol/mol] vs. ≥8.0% [≥63.9 mmol/mol]), history of myocardial infarction (MI), and age (<65 vs. ≥65 years) and were randomly assigned (3:3:3:1) to receive, in addition to their background metformin, 1 of 4 treatments at baseline: albiglutide 30 mg, sitagliptin 100 mg, glimepiride 2 mg, or placebo. Matching placebos for albiglutide, sitagliptin, and glimepiride were used to maintain blinding to treatment.[9] [An improvement on the reporting would be to explain why a 3:3:3:1 allocation was used.]*

*Explanation*: In terms of readability and understanding the design and rationale of a multi-arm trial, specification of the number of treatment groups is essential. Describing the allocation ratio offers insight and clarity, especially if an unequal allocation ratio is chosen, in which case an explanation is necessary.

Illustrating the structure and participant flow in a multi-arm trial will usually provide insight to the reader. An example demonstrating a trial structure and participant flow is shown in the eFigure in the Supplement.[9] Nevertheless, the presentation of the trial structure and participant flow in this example could be improved in terms of the labeling (eg, position of "Follow-up" in diagram A), the absence of 2 arrows leading from the randomization box, and the description of the information provided (eg, what is meant by "Terminated by sponsor" in diagram B?).

### Item 3b. CONSORT 2010: Important changes to methods after trial commencement (such as eligibility criteria), with reasons.

Extension for multi-arm trials: Details of any treatment groups added or dropped (if relevant) with reasons and/or changes to the allocation ratio.

*Example* (in which an arm was dropped).

*The original study was a multicentre, blinded, randomized, parallel-group trial in which patients were assigned to receive risperidone (Risperdal, Eisai), donepezil, or placebo for 12 weeks, after 4 weeks of psychosocial treatment. The target sample size was 285 people with Alzheimer's disease. Recruitment started in November 2003 but was suspended in March 2004, following the recommendation by the United Kingdom Committee for Safety of Medicines that risperidone and olanzapine not be used for the treatment of behavioral symptoms in dementia. The trial was restarted in July 2004 with a two-group design (donepezil and placebo), and recruitment ended in September 2005.[18]*

Box. Methodological Issues in Multi-Arm Randomized Trials

**Design**

Research Objectives

Trials with more than 2 treatment arms will generally either address a more complex question than a 2-arm trial or, more commonly, will attempt to address research questions about more than 1 intervention. Authors should explicitly define the objectives of a multi-arm trial, referring to all the arms of the study and prespecifying all planned comparisons of intervention groups to partly mitigate the effects of multiplicity and accusations of data dredging (ie, unplanned exploratory analyses).

Eligibility Criteria

In trials that involve multiple drugs, safety/toxicity profiles may reduce the pool of potential participants, adversely affecting recruitment and generalizability.

Patient and/or recruiting center characteristics or lack of equipoise or resources may preclude randomization to 1 of the groups.

A multi-arm trial could include 2 research treatments with contraindications but allow patients to be randomized to other arms. For example, everyone might not be suitable for a type of surgical procedure but may be able to contribute to an evaluation of a drug, so patients could be randomized as control vs surgery vs drug, control vs surgery (if not suitable for drug therapy), or control vs drug (if not suitable for surgery).

Sample Size

The sample size for a multi-arm phase 3 trial should depend upon the planned primary comparison(s). The sample size per group should be large enough that prespecified primary comparisons have adequate power.

Use of Placebos for Blinding

In the multi-arm design, blinding needs to ensure that none of the arms can be identified. If route of administration of 2 experimental drugs varies (eg, oral vs intravenous), blinding can become invasive, expensive, and an additional burden on participants. As the number of experimental drug arms increases, blinding may become more problematic.

3-Arm Trial: Placebo and Active Control

Three-arm trials that include an active control group as well as a placebo group can establish whether a failure to distinguish a test treatment effect from placebo implies ineffectiveness of the new test treatment or is simply the result of a trial that lacked the ability to identify an active drug. The comparison of placebo to the active control (standard drug) in such a design provides internal evidence of assay sensitivity (a property of a clinical trial defined as the ability of a trial to distinguish an effective treatment from a less effective or ineffective intervention). An unequal allocation ratio could be used to make the active groups larger than the placebo group to improve the precision of the active drug comparison. This allocation ratio may increase acceptability to participants and investigators because there is a lower probability of being allocated to placebo.[14]

**Conduct**

Interim Analysis and Stopping Guidelines

Many trials employ formal methods for interim monitoring and early stopping guidelines. These guidelines prompt consideration for recruitment to stop early for strong evidence of benefit or harm or, alternatively, futility. Multiple treatment arms add to the complexity of interpreting interim analyses in the context of early stopping guidelines. Depending on the type/structure of a multi-arm trial, an ethical dilemma may arise as a result of an interim analysis, such as if sufficiently strong evidence of a benefit of one of the treatment interventions vs the control is observed. If this intervention is considered a significant improvement over the control arm, recruitment into the control arm may have to be stopped, which may result in recruitment to the other treatment intervention arms stopping because of a lack of a concurrent control group. Because

the trial may be stopped if any of the treatment intervention-control comparisons cross an efficacy early stopping boundary, multiplicity adjustment is required for the efficacy boundaries.

**Analysis**

Analysis Strategy

If the main objective is to examine whether the interventions differ, but not how they differ, it would be appropriate to compare all groups at once using a single global test of significance. If the main objective is to examine a trend, a dose-response model should be used. More often, 2 or more specific comparisons are made between particular pairs or combinations of treatments. However, the number of possible comparisons can be considerable.

Multiple Treatments Comparisons

For a 3-arm trial (eg, treatments A, B, and C) there are several possible comparisons, including:

1. Comparing all 3 groups at once (A vs B vs C); a global test of unordered groups or a test for trend across ordered groups.

2. Comparing 1 group to the other 2 groups combined (A plus B vs C) and then the groups that were combined to each other (A vs B); A and B might be low and high doses of the same drug and the first comparison could be of treated vs untreated, followed by a comparison of the 2 treated groups, or A and B might be 2 antibiotics in the same class vs C as a member of a different class (note: the labeling in this example is arbitrary).

3. All pairwise comparisons: A vs B, A vs C, and B vs C.

4. Comparing A vs C and B vs C, but not A vs B; for example, comparing 2 treatments, separately, to the control but not comparing the 2 treatments to each other.

**Reporting and Interpretation**

Multi-arm trials often address complex and intricate questions concurrently and, as such, have a different focus than 2-arm trials. For example, following a prespecified comparison of all groups, the interpretation of a statistically significant global test is not straightforward. The investigators have evidence to reject the hypothesis that all the interventions were equally effective, but no clear indication of precisely where the differences lie. It is tempting, but incorrect, simply to use the observed data to draw more precise conclusions. It is incorrect, for example, to deduce that the intervention with the most favorable results is better than the others, because this question has not been examined explicitly.

Moreover, multiple pairwise comparisons may yield seemingly paradoxical results. For example, in a trial of 2 active interventions (A and B) vs placebo, it is possible to find that A is significantly better than the placebo but that B is not significantly different from either A or the placebo. It is also possible that no pairwise comparison is significant despite a significant global test. These problems are well known in agricultural research and other research areas in which formal multi-arm comparisons are common, but there is little experience of such issues in clinical research. Interpretation issues relating to the multiplicity of comparisons are of general relevance. Clinicians frequently find that the addition of a group to a trial enhances rather than diminishes the information gained.[15] In many such trials, interpretation of results adjusted for multiplicity frequently causes rather than solves interpretational problems. Yet, sometimes a particular analysis dictates adjustment for multiple comparisons; if those adjustments are indeed unsophisticated and liable to overcorrection, the authors should account for that in their interpretation.

Readers of a report of a multi-arm trial will expect a description of how the primary and secondary comparisons were handled, emanating from the multiple intervention groups. Most authors and readers are likely to bear in mind the number of analyses performed regardless of whether any formal adjustment is made.

*Example* (in which an arm was added).

*A total of 1493 patients with schizophrenia were recruited at 57 U.S. sites and randomly assigned to receive olanzapine (7.5 to 30 mg per day), perphenazine (8 to 32 mg per day), quetiapine (200 to 800 mg per day), or risperidone (1.5 to 6.0 mg per day) for up to 18 months. Ziprasidone (40 to 160 mg per day) was included after its approval by the Food and Drug Administration. The primary aim was to delineate differences in the overall effectiveness of these five treatments.[19]*

*Explanation*: If treatment arms are added or dropped, the number of participants available for an unbiased and valid comparison is affected (ie, only participants randomized concurrently should be compared). In the example above in which a treatment arm was discontinued, the allocation ratio went from 1:1:1 to 1:1 (evident from the participant flow diagram and results tables), so the probability of receiving one of the interventions changed from 0.33 to 0.50, but randomization continued with roughly equal probability of receiving either intervention. In the example of adding an arm, the allocation ratio was not explicitly ever mentioned.

This item relates to a conventional multi-arm trial and not to an adaptive design in which arms may be dropped using prespecified rules. Such designs offer greater efficiency while minimizing the number of participants that need to be randomized. Reporting guidelines for adaptive trials will be covered by the Adaptive Designs CONSORT Extension.[20]

**Item 7a. CONSORT 2010: How sample size was determined.**
Extension for multi-arm trials: Planned sample size with details of how it was determined for each primary comparison.

*Sample size calculations were based on the assumption that 34% of placebo-treated patients and 54-64% of tadalafil-treated patients (once daily and on demand) would achieve an IIEF-EF score [International Index of Erectile Function-Erectile Function] after DFW [drug-free washout]. A sample size of 412 randomised patients provided 84% power to detect a 20% difference in proportions in the two pairwise comparisons of tadalafil (once daily and on demand) versus placebo (20% drop-out rate assumed).[21]*

*Because a high degree of benefit would be needed to change routine clinical practice, we specified a 3.3% absolute reduction on the basis of estimated incidence in the control group of 11% (30% relative reduction; odds ratio [OR] 0.67). With 90% power and 2.5% significance level to account for the two comparisons, and allowing for an attrition rate of 15%, we needed to recruit 2,345 participants in each group (7,035 participants overall). Two comparisons of equal importance were tested in the trial: silver alloy catheters versus PTFE [polytetrafluoroethylene] catheters and nitrofural catheters versus PTFE catheters.[22]*

*Explanation*: The sample size for a multi-arm trial should correspond to the planned primary comparisons (item 2b). The approach to sample size is determined by the structure of the interventions being compared and the nature of the planned analyses (Box). When pairwise comparisons are planned, the sample size will usually be determined to give adequate power to evaluate each of the intended primary comparisons. If investigators deem that they need to adjust for multiple comparisons, the planned sample size may be inflated to account for that adjustment (item 7a).

**Item 12a. CONSORT 2010: Statistical methods used to compare groups for primary and secondary outcomes.**
Extension for multi-arm trials: Explicitly state if no adjustments for multiplicity were applied; if adjustments were made, state the method used.

*Examples* (in which adjustments were not made).

*The hypotheses were that the high-rate group, the delayed-therapy group, or both would have a reduced risk of a first occurrence of inappropriate therapy, as compared with the conventional-therapy group. The two trials were conducted in parallel, with inference made in each, and no adjustment for multiple comparisons was deemed appropriate.[23]*

*All P values are two-sided with no adjustment made for multiple comparisons.[24]*

*Examples* (in which adjustments were made).

*We assessed urinary tract infection outcomes with logistic regression and summarised findings as absolute percentage risk differences and ORs, both with 95% CIs calculated as 97·5% confidence intervals to adjust for the two comparisons. For the primary analysis, p=0·025 was regarded as significant.[22]*

*For both [Visual Analogue Scale] VAS-immediate pain ratings and pressure data, if the Shapiro-Wilk normality test was passed, repeated measures one-way ANOVA [Analysis of variance] with Bonferroni correction post-hoc pairwise comparisons was conducted to explore any significant difference (P<0.05) between the test conditions.[25]*

*... we calculated that we would need to enroll 810 patients in each group for the study to have 90% power to show the superiority of apixaban over placebo, at a two-sided alpha level of 0.05, with the use of the Hochberg multiple-testing method.[26]*

*Explanation*: In general, multi-arm trial analysis strategies may have 2 broad objectives. First, investigators examine variation in efficacy of several interventions, which can be addressed by an overall analysis comparing all groups at once. Such an analysis is unlikely to be fully satisfactory because it will not indicate the areas of differences. Second, 2 or more specific pairwise comparisons can be made between particular treatments, as described above. In a particular trial, both types of analyses may be performed. One strategy (commonly recommended in agricultural analyses and other experiments) is to first perform a global statistical test across all groups, and only to proceed to paired comparisons if the global test is statistically significant. This strategy does not seem especially desirable for the analysis of clinical trials, which require a more focused approach to the evaluation of treatment comparisons.

Two further complications may be present. First, 2 (or more) of the treatments may be different doses or durations of the same drug or intervention. In such cases, it may be of most interest to examine whether there is a dose-response relation rather than simply testing the significance of differences between pairs of treatments. Second, 2 of the groups may receive variants of the same basic intervention. For example, they may receive the same drug either orally or intravenously. Investigators might first compare these groups combined vs the comparison group (usually placebo or standard treatment) before considering whether the 2 variants might differ. Groups receiving different doses may

also sometimes be considered in this way. When such an analysis is planned, researchers may sometimes consider that the groups should be allocated in a 1:1:2 ratio to maximize the power of the first comparison. Statistical adjustment for multiple comparisons invokes debate among methodologists, and there is no consensus. While some would use such an adjustment, others would never apply adjustments.[27,28]Investigators may avoid multiplicity problems with analytical approaches. Some examples include:

- Using a single global test of significance across comparison groups (eg, comparing A vs B vs C in a 3-arm trial) and avoiding multiple comparisons. Of note, a single global test across all the treatments is of limited use.[29]
- Modeling a dose-response relationship and eliminating multiple comparisons.[30]
- Using a prioritized sequence of tests. For example, investigators might decide upon the new 300-mg antibiotic vs standard treatment as the priority test and, if that comparison is statistically significant, continue to the 200-mg antibiotic vs standard treatment comparison. A prioritized sequence of tests addresses multiplicity without adjustments.[31]
- Not making adjustments for multiplicity while transparently reporting all comparisons made. Many multi-arm trials are designed for direct comparison of unrelated treatments with a control arm, such as comparing A vs C and B vs C in a 3-arm trial. Adjustments for multiple comparisons generally need not play a role in such multi-arm trials.[2,15,32,33]

Sometimes formal adjustments for multiplicity are unavoidable; some regulators, such as the European Medicines Agency, require such adjustments. As stated in the European Medicines Agency's guideline on multiplicity issues in clinical trials, "as a general rule it can be stated that control of the study-wise type I error is a minimal prerequisite for confirmatory claims."[34] However, even when adjustment becomes appropriate, implementation becomes problematic. Bonferroni adjustments are often recommended, usually because of their simplicity. However, other adjustment strategies sometimes perform better on the overall control of the type-1 error rate (usually called the family-wise type-1-error [FWER]),[32,35-37] while performing worse on the probability of more than 1 false positive.[32] The adjustments frequently provide overcorrection for multiplicity, especially the Bonferroni adjustment. This approach becomes overly conservative as the correlation among the comparisons becomes higher. Other approaches, including the Holm, Hochberg, Dunnett test, and adjusted Hochberg mehod, have been compared to the Bonferroni approach.[32] All methods appear less conservative than the Bonferroni.

## Results

### Item 14a. CONSORT 2010: Dates defining the periods of recruitment and follow-up.

Extension for multi-arm trials: If periods of recruitment and follow-up are different across treatment groups (eg, groups were added or dropped), the periods of recruitment and follow-up, reason(s) for the differences, and any statistical implications should be described.

Methods (Study Setting and Design): The study was conducted between January 2001 and December 2004 at 57 clinical sites in the United States (16 university clinics, 10 state mental health agencies, 7 Veterans Affairs medical centers, 6 private non-profit agencies, 4 private-practice sites, and 14 mixed-system sites). Patients were initially randomly assigned to receive olanzapine, perphenazine, quetiapine, or risperidone under double-blind conditions and followed for up to 18 months or until treatment was discontinued for any reason (phase 1). (Ziprasidone was approved for use by the Food and Drug Administration [FDA] after the study began and was added to the study in January 2002 in the form of an identical-appearing capsule containing 40 mg).

Methods (Statistical Analysis): ... Ziprasidone was added to the trial after approximately 40 percent of the patients had been enrolled... and comparisons involving the ziprasidone group were limited to the cohort of patients who underwent randomization after ziprasidone was added (the ziprasidone cohort). In general, the trial had a statistical power of 85 percent to identify an absolute difference of 12 percent in the rates of discontinuation between two atypical agents; however, it had a statistical power... of 58 percent for comparisons involving ziprasidone... The overall difference among the olanzapine, quetiapine, risperidone, and perphenazine groups was evaluated with the use of a test with 3 degrees of freedom (df). If the difference was significant at a P value of less than 0.05, the three atypical-drug groups were compared with each other by means of step-down or closed testing, with a P value of less than 0.05 considered to indicate statistical significance... The ziprasidone group was directly compared with the other three atypical-drug groups and the perphenazine group within the ziprasidone cohort by means of a Hochberg adjustment for four pairwise comparisons. The smallest resulting P value was compared with a value of 0.013 (0.05÷4). [reiterated in a footnote to Table 2 and Figure 2 legend relating to Outcome Measures of Effectiveness in the Intention-To-Treat (ITT) population]

Results (Discontinuation of Treatment): ... Within the cohort of 889 patients who underwent randomization after ziprasidone was added to the trial, those receiving olanzapine had a longer interval before discontinuing treatment for any cause than did those in the ziprasidone group (hazard ratio, 0.76; P=0.028). However, this difference was not significant after adjustment for multiple comparisons (required P value, ≤0.013).[19]

Explanation: Incorporating an emerging therapy as a new randomization group in a clinical trial that is open to recruitment would be desirable to researchers, regulators, and patients to ensure that the trial remains current, new treatments are evaluated as quickly as possible, and the time and cost for determining optimal therapies is minimized.[38] Numerous methodological and statistical implications should be considered. These implications include (1) family-wise error rate control due to stage effects and multiplicity, (2) that only concurrent control group data are used for an unbiased comparison with the added arm(s),[39] (3) statistical power (comparison with concurrent control group data will require adequate power), (4) the allocation ratio and/or length of recruitment into each group (improved efficiency could be realized by adjusting the total number of participants required and time spent recruiting to answer the primary hypotheses), (5) potential changes to the control group (it is possible that the existing control group may be shown to be inferior and, therefore, it is theoretically possible that the control group may have to be changed), and (6) logistical considerations (eg, extra funding, the time taken for all necessary approvals/amendments,

sourcing drug, updating trial randomization and clinical database systems, possible effect on blinding, trial oversight, recruitment).[38] The extent to which these implications need to be considered depends upon the nature and structure of the trial. There is potential for overlap with the CONSORT extension for adaptive designs.[20]

If recruitment into more than 1 treatment group in a multi-arm trial is stopped prematurely, it is important to include the reasons why, because those reasons may differ. In addition, regarding standard CONSORT item 15 (ie, a table showing baseline demographic and clinical characteristics for each group), in a situation in which recruitment to all treatment groups is not contemporaneous, a single table or multiple baseline tables could be used. Authors must clearly state which participants are included in which comparisons for each group.

### Item 17a. CONSORT 2010: For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).

Extension for multi-arm trials: Results for each prespecified comparison of treatment groups.

*Primary Outcomes – At 6 months, the AVVQ [Aberdeen Varicose Veins Questionnaire] score in the foam group was significantly higher (indicating a worse disease-specific quality of life) than that in the surgery group, but the difference was moderate (effect size, −1.74; 95% confidence interval [CI], −2.97 to −0.50; P=0.006). The improvement in the AVVQ score in the laser group did not differ significantly from that in the surgery group. There were no significant differences between the groups in the EQ-5D score [a standardized instrument for measuring generic health status] or the SF-36 [Short Form Health Survey] physical component score. For the post hoc analysis of treatment with laser versus foam, the only significant difference was in the SF-36 mental component score, which was slightly higher (better generic quality of life) in the laser group than in the foam group (effect size, 1.54; 95% CI, 0.01 to 3.06; P=0.048)… Secondary Outcomes – Quality of Life. At 6 weeks, significant between-group differences (P<0.005) included a lower AVVQ score (indicating a better disease-specific quality of life) in the surgery group than in the foam group (effect size, −2.3; 95% CI, −3.7 to −0.9) and lower SF-36 scores (indicating a worse generic quality of life) in the surgery group than in the laser group for the domains of bodily pain (effect size, −2.7; 95% CI, −4.4 to −0.9), vitality (effect size, −2.3; 95% CI, −3.9 to −0.8), role limitations due to emotional health (effect size, −2.4; 95% CI, −4.0 to −0.8), and role limitations due to physical health (effect size, −3.5; 95% CI, −5.2 to −1.8). These four SF-36 domain scores did not differ significantly (with P<0.005 considered to indicate statistical significance) between groups at 6 months. For the post hoc comparisons of laser treatment vs foam treatment, only the EQ-5D score was significantly lower (indicating a worse generic quality of life) in the foam group at 6 weeks (0.044; 95% CI, 0.014 to 0.074).[40]*

*Explanation*: Investigators should plan the comparisons intended, document them in the protocol and statistical analysis plan, and report them all in the trial report with appropriate interpretations. If intervention groups have been added or dropped during the trial, it is important that the analysis addresses the implications of doing so. If investigators employed measures to control the overall sig-

nificance level (eg, if they conducted a single global test of significance across comparison groups, modeled a dose-response relationship, or used a prioritized sequence of tests), those details should be reported. If investigators conducted an analysis that dictated formal adjustments for multiplicity, those methods and limitations should be reported. As discussed previously (item 12a), many multi-arm trials will not employ formal adjustments for multiplicity. In those cases, investigators should still transparently report all comparisons undertaken, planned and unplanned, and provide appropriate interpretations of the results.

### Discussion

**Item 20. CONSORT 2010: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.**
While no specific extension to the standard CONSORT item is recommended here, authors should address the strengths and limitations of multi-arm trials with regard to issues detailed in the Box.

## Discussion

Multi-arm trials require careful thought and planning. They offer the opportunity to address more than 1 research question, may accelerate the evaluation of new interventions, and can facilitate head-to-head comparisons with competing treatment options, potentially resulting in patient benefit while optimizing the use of resources. Multi-arm trials may be more appealing than trials with only 2 arms to participants and clinicians because typically there is an increased probability of receiving an experimental intervention rather than standard care. However, investigators should always be mindful that the efficiency advantages of multi-arm trials and the opportunity to evaluate more interventions in a shorter time are contingent upon recruiting and collating outcomes on the requisite number of participants.

Multi-arm randomized trials are common, and it is important that reports of these trials include information on features specific to the design to allow readers to make an accurate assessment of the conduct of the trial and interpretation of the results. Transparent and complete reporting is an essential prerequisite for reproducibility. Good reporting also facilitates the identification and inclusion of multi-arm trials in systematic reviews. However, multi-arm trials, especially trials with more than 3 treatment arms, are challenging to design and analyze.

This Special Communication provides a proposed extension to the widely adopted CONSORT 2010 Statement to enable the full and accurate reporting of multi-arm randomized trials. Such trials require clear objectives and hypotheses referring to all of the treatment arms and identification of the primary comparisons being made. The sample size should be prespecified and the issue of adjustment for multiple testing should at least be acknowledged. If periods of recruitment and follow-up are different across treatment groups (eg, groups were added or dropped), the periods and reasons for differences should be reported, and any statistical implications should be addressed.

Multiplicity adjustment for multiple comparisons among groups in a multi-arm randomized trial remains a challenging issue. Many multi-arm trials are conducted for efficiency reasons. They compare distinct treatments/interventions against a single

control group, which could easily have been done in multiple separate trials rather than a single multi-arm trial. For a multi-arm trial design in which several experimental interventions share a control arm, the trial is focused on evaluating the research question for each intervention separately. The interpretation of the results of one comparison ordinarily has no direct bearing on the interpretation of the others. Many trialists/methodologists argue that multiplicity adjustments are not necessary in such instances because such adjustments would not be necessary if the interventions were compared in separate trials.[2,15,32,33,41,42] Some multi-arm trials evaluate several different doses of the same agent against a control group, which represents related comparisons. In such situations, trialists and methodologists tend to recommend multiplicity adjustments.[2,32,33,37] An example of this situation occurs with certain decision-making criteria in submissions to a regulatory agency for drug approval. If the sponsor specifies more than 1 treatment comparison and proposes to claim a treatment effect if 1 or more of the doses are statistically significant, most trialists and methodologists suggest an adjustment for multiplicity.[2,15,32,33,41,42] But sweeping declarations of always or never needing to adjust for multiple testing should be ignored; the decision regarding adjustment depends on the objectives, design, and analysis.

Some multi-arm trials may also have other special features, such as being crossover, cluster, or factorial trials. For such trials, the specific recommendations for all such types of trial will be relevant. Use of the CONSORT Statement for the reporting of 2-group parallel trials has been shown to be associated with improved quality of reporting.[43] The routine use of this proposed extension to the CONSORT Statement is intended to promote similar improvements.

The CONSORT Group will continue to monitor and revise its recommendations and is developing checklists and flow diagrams to help improve the quality of reporting of clinical trials of various designs. Other similar extensions and updates are in preparation, and the most up-to-date versions of all CONSORT recommendations can be found on the CONSORT website (http://www.consort-statement.org).

## Conclusions

This extension of the CONSORT 2010 Statement provides specific guidance for the reporting of multi-arm parallel-group randomized clinical trials and should help to provide greater transparency and accuracy in the reporting of these types of clinical trials.

## REFERENCES

**1**. Parmar MK, Carpenter J, Sydes MR. More multiarm randomised trials of superiority are needed. *Lancet*. 2014;384(9940):283-284. doi:10.1016/S0140-6736(14)61122-3

**2**. Freidlin B, Korn EL, Gray R, Martin A. Multi-arm clinical trials of new agents: some design

considerations. *Clin Cancer Res*. 2008;14(14):4368-4371. doi:10.1158/1078-0432.CCR-08-0325

**3**. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272(2):122-124. doi:10.1001/jama.1994.03520020048013

**4**. Odutayo A, Emdin CA, Hsiao AJ, et al. Association between trial registration and positive study findings: cross sectional study (Epidemiological Study of Randomized Trials-ESORT). *BMJ*. 2017;356:j917. doi:10.1136/bmj.j917

**5**. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869. doi:10.1136/bmj.c869

**6**. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. doi:10.1136/bmj.c332

**7**. Hopewell S, Clarke M, Moher D, et al; CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med*. 2008;5(1):e20. doi:10.1371/journal.pmed.0050020

**8**. Ioannidis JP, Evans SJ, Gøtzsche PC, et al; CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141(10):781-788. doi:10.7326/0003-4819-141-10-200411160-00009

**9**. Ahrén B, Johnson SL, Stewart M, et al; HARMONY 3 Study Group. HARMONY 3: 104-week randomized, double-blind, placebo- and active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin. *Diabetes Care*. 2014;37(8):2141-2148. doi:10.2337/dc14-0024

**10**. Agar MR, Lawlor PG, Quinn S, et al. Efficacy of oral risperidone, haloperidol, or placebo for

symptoms of delirium among patients in palliative care: a randomized clinical trial. *JAMA Intern Med*. 2017;177(1):34-42. doi:10.1001/jamainternmed.2016.7491

11. Connick P, De Angelis F, Parker RA, et al; UK Multiple Sclerosis Society Clinical Trials Network. Multiple Sclerosis-Secondary Progressive Multi-Arm Randomisation Trial (MS-SMART): a multiarm phase IIb randomised, double-blind, placebo-controlled clinical trial comparing the efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis. *BMJ Open*. 2018;8(8):e021944. doi:10.1136/bmjopen-2018-021944

12. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof*. 2002;25 (1):38-64.

13. Geddes JR, Goodwin GM, Rendell J, et al; BALANCE investigators and collaborators. Lithium plus valproate combination therapy versus monotherapy for relapse prevention in bipolar I disorder (BALANCE): a randomised open-label trial. *Lancet*. 2010;375(9712):385-395. doi:10.1016/S0140-6736(09)61828-6

14. European Medicines Agency. *ICH Topic E 10: Choice of Control Group in Clinical Trials*. Canary Wharf, London, United Kingdom: European Medicines Agency; 2001. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf. Accessed February 14, 2018.

15. Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. *Lancet*. 2005;365(9470):1591-1595. doi:10.1016/S0140-6736(05)66461-6

16. Foa EB, McLean CP, Zang Y, et al; STRONG STAR Consortium. Effect of prolonged exposure therapy delivered over 2 weeks vs 8 weeks vs present-centered therapy on PTSD symptom severity in military personnel: a randomized clinical trial. *JAMA*. 2018;319(4):354-364. doi:10.1001/jama.2017.21242

17. Gray R, Ives N, Rick C, et al; PD Med Collaborative Group. Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): a large, open-label, pragmatic randomised trial. *Lancet*. 2014;384(9949):1196-1205. doi:10.1016/S0140-6736(14)60683-8

18. Howard RJ, Juszczak E, Ballard CG, et al; CALM-AD Trial Group. Donepezil for the treatment of agitation in Alzheimer's disease. *N Engl J Med*. 2007;357(14):1382-1392. doi:10.1056/NEJMoa066583

19. Lieberman JA, Stroup TS, McEvoy JP, et al; Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353(12):1209-1223. doi:10.1056/NEJMoa051688

20. Dimario M, Todd S, Julious S, et al. Adaptive designs CONSORT Extension (ACE) Project Protocol: version 2.3. http://www.equator-network.org/wp-content/uploads/2017/12/ACE-Project-Protocol-v2.3.pdf. Published 2016. Accessed February 14, 2018.

21. Montorsi F, Brock G, Stolzenburg JU, et al. Effects of tadalafil treatment on erectile function recovery following bilateral nerve-sparing radical prostatectomy: a randomised placebo-controlled study (REACTT). *Eur Urol*. 2014;65(3):587-596. doi:10.1016/j.eururo.2013.09.051

22. Pickard R, Lam T, MacLennan G, et al. Antimicrobial catheters for reduction of symptomatic urinary tract infection in adults requiring short-term catheterisation in hospital: a multicentre randomised controlled trial. *Lancet*. 2012;380(9857):1927-1935. doi:10.1016/S0140-6736(12)61380-4

23. Moss AJ, Schuger C, Beck CA, et al; MADIT-RIT Trial Investigators. Reduction in inappropriate therapy and mortality through ICD programming. *N Engl J Med*. 2012;367(24):2275-2283. doi:10.1056/NEJMoa1211107

24. Ndibazza J, Mpairwe H, Webb EL, et al. Impact of anthelminthic treatment in pregnancy and childhood on immunisations, infections and eczema in childhood: a randomised controlled trial. *PLoS One*. 2012;7(12):e50325. doi:10.1371/journal.pone.0050325

25. Fong DT, Pang KY, Chung MM, Hung AS, Chan KM. Evaluation of combined prescription of rocker sole shoes and custom-made foot orthoses for the treatment of plantar fasciitis. *Clin Biomech (Bristol, Avon)*. 2012;27(10):1072-1077. doi:10.1016/j.clinbiomech.2012.08.003

26. Agnelli G, Buller HR, Cohen A, et al; AMPLIFY-EXT Investigators. Apixaban for extended treatment of venous thromboembolism. *N Engl J Med*. 2013;368(8):699-708. doi:10.1056/NEJMoa1207541

27. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ*. 1998;316(7139):1236-1238. doi:10.1136/bmj.316.7139.1236

28. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1(1):43-46. doi:10.1097/00001648-199001000-00010

29. Pocock SJ. *Clinical Trials: A Practical Approach*. Chichester, UK: John Wiley & Sons Ltd; 1983.

30. Senn S. *Statistical Issues in Drug Development*. Chichester, UK: John Wiley & Sons Ltd; 1997.

31. Bauer P, Chi G, Geller N, et al. Industry, government, and academic panel discussion on multiple comparisons in a "real" phase three clinical trial. *J Biopharm Stat*. 2003;13(4):691-701. doi:10.1081/BIP-120024203

32. Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Stat Methods Med Res*. 2018;27(5):1513-1530. doi:10.1177/0962280216664759

33. Wason JM, Stecher L, Mander AP. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? *Trials*. 2014;15:364. doi:10.1186/1745-6215-15-364

34. European Medicines Agency. *Guideline on Multiplicity Issues in Clinical Trials* [draft]. Canary Wharf, London, United Kingdom: European Medicines Agency; 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf. Accessed February 14, 2018.

35. Sankoh AJ, D'Agostino RB Sr, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med*. 2003;22(20):3133-3150. doi:10.1002/sim.1557

36. Hsu JC. *Multiple Comparisons: Theory and Methods*. New York, NY: Chapman & Hall; 1996.

37. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. *Control Clin Trials*. 2000;21(6):527-539. doi:10.1016/S0197-2456(00)00106-9

38. Cohen DR, Todd S, Gregory WM, Brown JM. Adding a treatment arm to an ongoing clinical trial: a review of methodology and practice. *Trials*. 2015;16:179. doi:10.1186/s13063-015-0697-y

39. Altman DG. Avoiding bias in trials in which allocation ratio is varied. *J R Soc Med*. 2018;111(4):143-144. doi:10.1177/0141076818764320

40. Brittenden J, Cotton SC, Elders A, et al. A randomized trial comparing treatments for varicose veins. *N Engl J Med*. 2014;371(13):1218-1227. doi:10.1056/NEJMoa1400781

41. Duncan DB. Multiple range and multiple F tests. *Biometrics*. 1955;11:1-42. doi:10.2307/3001478

42. Cook RJF. V. T. Multiplicity considerations in the design and analysis of clinical trials. *J Royal Stat Soc A Stat Soc*. 1996;159(1):93-110. doi:10.2307/2983471

43. Turner L, Shamseer L, Altman DG, et al. Consolidated Standards of Reporting Trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11:MR000030.